Ab-Initio Fragment Method for Calculating Molecular X-ray Diffraction

Ab-Initio Fragment Method for Calculating Molecular X-ray Diffraction

Thomas Northey[†,‡] and Adam Kirrander[*,†]

†EaStCHEM, School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom

‡Current address: School of Natural Sciences and Environment, Newcastle University, Newcastle, United Kingdom

E-mail: Adam.Kirrander@ed.ac.uk
Phone: +44 (0)131 6504716

Abstract

A fragment-based approach for the prediction of elastic x-ray scattering is presented. The total diffraction pattern is assembled from anisotropic form factors calculated for individual molecular fragments, optionally including corrections for pair-wise interactions between fragments. The approach is evaluated against full *ab-initio* scattering calculations in the peptide diphenylalanine, and the optimal selection of fragments is examined in the ethanol molecule. The approach is found to improve significantly on the independent atom model, while remaining conceptually simple and computationally efficient. It is expected to be particularly useful for macromolecules with repeated subunits, such as peptides, proteins, DNA or RNA, and other polymers where it is straightforward to define appropriate fragments.

1 Introduction

With the arrival of X-ray Free-Electron Lasers (XFELs)[1–4] molecules can be probed with bright ultrashort x-ray pulses,[5] making it possible to track changes in molecular structure on femtosecond timescales using pump-probe x-ray scattering in the gas-phase[6–10] and in solution.[11,12] Despite the small scattering cross sections for x-rays, the experiments do not require crystalline samples thanks to the large number of x-ray photons. At least in principle, sub-Ångström resolution[13] can be achieved and, importantly, the experiments provide information that is complementary to ultrafast spectroscopies.[14–17] We also note that ultrafast electron diffraction is a closely related technique that uses high-energy electrons instead of photons.[18–21]

XFELs are also transforming structural biology. A major bottleneck in x-ray crystallography is to grow crystals of biomolecules, but the high brightness of XFELs makes it potentially possible to dispense with crystals altogether.[5] In serial femtosecond x-ray crystallography,[22–25] a liquid jet containing 100,000s of microcrystals, which are too small for conventional crystallography, intersects the femtosecond x-ray pulses.[26,27] Many snapshots of microcrystals in different orientations are recorded, and computer algorithms are used to extract molecular structure from the data. Ongoing reductions in the necessary size of the protein crystals[28] and new correlation-based diffraction methods[29] point towards an era of single-molecule structure determination.

Although attempts at greater sophistication in the analysis of high-resolution x-ray diffraction already exist,[30] x-ray structure determination is dominated by the independent atom model (IAM).[31] IAM is a simple and efficient approximation, but fails to capture the redistribution of electrons that is a key aspect of e.g. chemical bonding. One would ideally wish to capture subtle

changes in molecular bonding or electronic states, something that requires better accuracy than provided by IAM. The quest for more sophisticated analysis is further fuelled by the expectation of higher quality data. The partial or full alignment achieved in gas-phase studies[9,32] improves the quality of experimental observations, especially when statistical noise is minimized by the preparation of quantum mechanically identical states. In structural biology one expects that the volume of x-ray scattering data will increase since single-molecule diffraction gives a continuous signal that is no longer restricted to discrete Bragg peaks. Furthermore, the data is anticipated to be of higher quality since the scattering occurs before the sample has accumulated radiation damage[26] and the ultrashort duration of the XFEL pulses will mean that molecules will be frozen in instantaneous conformations rather than statistically averaged over thousands of microstates.

Based on our own interest in dynamics and ultrafast imaging, we have in recent years developed methods to calculate elastic,[33] inelastic,[34] and total[35] x-ray scattering directly from *ab initio* electronic structure calculations. We have further considered the effect of electronic,[33,36] vibrational and rotational[37–39] states on the scattering, have extensively investigated the intersection of quantum dynamics and scattering,[40–43] and have been involved in interpreting new experiments[6,7,9,10,14,17,44,45] performed at the LCLS at Stanford in California.[46]

Our starting point is the *ab initio* x-ray diffraction (AIXRD) method[33] which allows calculation of the molecular scattering (diffraction) factors directly from electronic structure calculations such as Hartree-Fock (HF), density functional theory (DFT), or various multiconfigurational methods such as complete active-space self-consistent field (CASSCF). We present a full derivation of AIXRD for Gaussian-type basis set. However, AIXRD is a computationally expensive approach compared to IAM, especially for large molecules such as proteins.

Fragment-based electronic structure methods have opened the way for quantum mechanical treatment of large molecules, such as biomolecules,[47] and make possible nearly linear scaling of calculations of large molecular systems, such as water clusters, proteins, and DNA.[48] Inspired by the divide-and-conquer approach of fragment molecular orbital (FMO) theory[48,49] and similar theories such as subsystem DFT,[50] and the division of the molecule into (atomic) subsystems by IAM, we explore a similar approach for elastic scattering. Combining AIXRD with a fragment-based method that divides the molecule into polyatomic subunits allows efficient calculation of scattering, and avoids the exponential scaling implicit in AIXRD.

2 Theory

2.1 X-ray diffraction

Structure determination using x-ray diffraction relies on the direct relationship between the electron density of a molecule and its (elastic) scattering signal. The intensity of elastic scattering is proportional to the absolute square of the molecular scattering form-factor which is defined as the Fourier transform of the electron density,

$$f(\mathbf{q}) = \int \rho(\mathbf{r};\mathbf{R})\, e^{i\mathbf{q}\cdot\mathbf{r}} \mathrm{d}\mathbf{r}, \tag{1}$$

for scattering vector $\mathbf{q}$, electronic coordinates $\mathbf{r}$, and the electron density $\rho(\mathbf{r};\mathbf{R})$ which depends parametrically on the nuclear coordinates $\mathbf{R}$. The scattering vector is defined by,

$$\mathbf{q} \equiv \mathbf{k}_i - \mathbf{k}_f, \tag{2}$$

for incident and final wavevectors, $\mathbf{k}_i$ and $\mathbf{k}_f$ respectively. The scattered radiation is detected at a detector far removed from the scattering source, i.e. in the far-field limit. For elastic scattering, $k_0{=}|\mathbf{k}_i|{=}|\mathbf{k}_f|$, the deflection angle θ of the scattered wavevector is related to $q{=}|\mathbf{q}|$ via,

$$q(\theta) = 2k_0 \sin\frac{\theta}{2}, \tag{3}$$

where $k_0{=}2\pi/\lambda$ for x-ray wavelength λ. For anisotropic scattering in the case of e.g. aligned molecules, an azimuthal angle ϕ would indicate anti-clockwise rotation around the centre of the detector.

The elastic scattering can be calculated directly from electronic structure calculation. In *ab initio*

x-ray diffraction (AIXRD) methods, the electron density is taken directly from the molecular wave function. In Kohn-Sham and Hartree-Fock theory the total electron density is a sum of molecular orbital (MO) electron densities,

$$\rho(\mathbf{r}) = \sum_{j=1}^{N_{\mathrm{MO}}} a_j |\phi_j(\mathbf{r})|^2, \tag{4}$$

where $\mathbf{r}$ represents the electronic coordinates, occupancies $a_j \in \{0, 1, 2\}$ (occupancy can be 1 in an unrestricted approach), and there are N_{MO} occupied MOs. Expressed in a contracted Gaussian basis set, as commonly used in molecular quantum chemistry calculations, this is,

$$\rho(\mathbf{r}) = \sum_{j=1}^{N_{\mathrm{MO}}} a_j \left| \sum_{k=1}^{N_{\mathrm{BF}}} M_k \sum_{i=1}^{N_{\mathrm{g}}^{(k)}} c_i g_i(\mathbf{r} - \mathbf{r}_i) \right|^2, \tag{5}$$

for N_{BF} basis functions (or contractions) with M_k orbital coefficients, and $N_{\mathrm{g}}^{(k)}$ Gaussian type-orbitals (GTOs) per kth contraction, each with fixed basis set cooefficient c_i and centered at coordinate $\mathbf{r}_i = (x_i, y_i, z_i)$,

$$g(\mathbf{r} - \mathbf{r}_i) = A \prod_{r=x,y,z} (r - r_i)^{l_r} e^{-\gamma(r-r_i)^2}, \tag{6}$$

with exponent γ and Cartesian orbital angular momentum $L = l_x + l_y + l_z$, where $l_r \in \mathbb{N}$, and normalisation constant is defined as,

$$A = \left(\frac{2}{\pi}\right)^{3/4} \frac{2^{(l_x+l_y+l_z)} \gamma^{(2l_x+2l_y+2l_z+3)/4}}{[(2l_x-1)!!(2l_y-1)!!(2l_z-1)!!]^{1/2}}, \tag{7}$$

for $(2l-1)!!=1\cdot3\cdot5\cdots(2l-1)$. Note that $\mathbb{N}$ denotes the set of integer numbers equal to and greater than zero.

The electron density in Eq. (5) can be Fourier transformed analytically.[33] Briefly, and not shown explicitly in Ref. 33, inserting Eq. (5) into Eq. (1), gives a sum of Fourier integrals of Gaussian products for each Cartesian coordinate, with different angular momentum numbers, l_i and $l_j \in \mathbb{N}$. Using the binomial theorem (twice), these integrals have the solution,

$$\begin{aligned} &\int g_i(r - r_i) g_j(r - r_j)\, e^{\imath q r} \mathrm{d}r = \\ &\sum_{m=0}^{l_i} \sum_{n=0}^{l_j} \sum_{p=0}^{m+n} \binom{l_i}{m} \binom{l_j}{n} \binom{m+n}{p} (-r_i)^{l_i-m} (-r_j)^{l_j-n} \\ &\times (r_0 + \imath q/2\alpha)^{m+n-p}\, e^{-\frac{q^2}{4\alpha} + \imath q r_0} H(p, \alpha), \end{aligned} \tag{8}$$

where from the Gaussian product theorem,[51] the centre of the new Gaussian formed by $g_i(r - r_i)g_j(r - r_j)$ is $r_0 = (\gamma_i r_i + \gamma_j r_j)/(\gamma_i + \gamma_j)$, and its exponent is $\alpha = \gamma_i + \gamma_j$, The remaining integral is,

$$H(p, \alpha) = \int u^p\, e^{-\alpha u^2} \mathrm{d}u, \tag{9}$$

for $u{=}r{-}r_0{+}\imath q/2\alpha$, which for odd values of $p \in \mathbb{N}$ is zero, and the even case has an analytical solution, thus,

$$H(p, \alpha) = \begin{cases} \frac{(p-1)!!}{(2\alpha)^{l/2}} \sqrt{\frac{\pi}{\alpha}}, & \text{if } p \in 2\mathbb{N}. \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

which gives a solution to Eq. (8) in the form of a reciprocal space polynomial multiplied by a Gaussian, as expected for a Gaussian Fourier transform problem.[38] Thus, the overall solution to Eq. (1) is a sum of $N_{\mathrm{MO}}(N_{\mathrm{GTO}})^2$ instances of Eq. (8), where $N_{\mathrm{GTO}}{=}N_{\mathrm{BF}}N_{\mathrm{g}}^{(k)}$ is the number of GTOs per MO (*cf.* Eq. (5)), with the appropriate orbital coefficients, M_k and $M_{k'}$, normalisation constants A_i and A_j, and basis set coefficients c_i and c_j, applied.

2.2 Fragment-based approaches

The exponential scaling inherent in the *ab initio* (AIXRD) approach can be overcome by using a fragment-based method to coarse-grain the calculations. A natural decomposition, originally proposed by Debye,[52,53] is to consider the molecule as a collection of isolated atoms that scatter independently. The independent atom model (IAM) thus approximates the molecular scattering form factor as a sum of atomic scattering factors,

$$f(\mathbf{q}) = \sum_{j}^{N_{\mathrm{at}}} f_j^0(q)\, e^{\imath \mathbf{q} \mathbf{R}_j}, \tag{11}$$

where N_{at} is the number of atoms, $f_j^0(q)$ the isotropic scattering form factor[54–56] for the *jth* atom (calculated for the atom in isolation), and $\mathbf{R}_j$ are its atomic coordinates. Conveniently, the atomic scattering form factors have been parametrized as a sequence of Gaussian functions,

$$f^0(q) = \sum_{k=1}^{4} \left[a_k \, e^{-b_k (q/4\pi)^2} \right] + c, \tag{12}$$

with the values a_k, b_k, and c tabulated for each atom and many atomic ions in the International Tables for Crystallography.[56] The independent atom model is extremely efficient computationally, but suffers from well-established drawbacks, especially in molecules with second row atoms that have comparatively few electrons and where the distortion of the electron density by molecular bonding is significant.[33,37]

We propose the independent fragment model (IFM) as a midway approximation in terms of cost and accuracy between full *ab initio* x-ray diffraction calculations (AIXRD)[33] and IAM. The molecule is decomposed into N_{f} fragments, which are the polyatomic chemical building blocks of the molecule. The scattering is calculated for each fragment individually using AIXRD, resulting in a set of fragment scattering factors $\{f_j^{IFM}(\mathbf{q})\}$, which are then summed together,

$$f_{\text{IFM}}(\mathbf{q}) = \sum_{j=1}^{N_{\text{f}}} f_j^{IFM}(\mathbf{q}), \tag{13}$$

to give the total molecular scattering factor $f_{\text{IFM}}(\mathbf{q})$. If the fragment form factors are not directly calculated in the appropriate molecular coordinates (for instance if they are taken from a fragment library of pre-calculated form factors) then the form factors $f_j^{IFM}(\mathbf{q})$ must be rotated to match their orientation in the molecule and given overall translational phase factors to match their position in the molecule. Currently, the form factors are calculated *on-the-fly* in the actual coordinates in the molecule, thus bypassing any need to translate and rotate the fragment into position.

Further corrections can be attempted by considering dimers defined as pairs of fragments, in order to account for pair-wise fragment interactions.

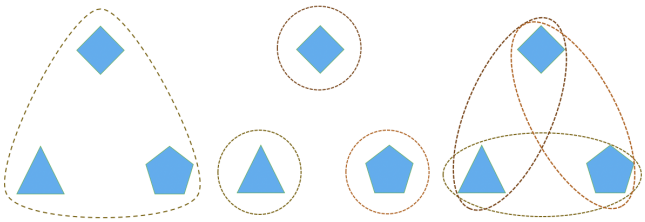

Figure 1: Schematic comparison of possible decompositions of a molecule into fragments. (Left) Whole molecule, no fragments. (Center) Three individual fragments. (Right) Three dimers, each consisting of a pair of fragments.

We denote this the dimer fragment model (DFM). The scattering from dimer fragments, given by the dimer form factors $\{f_j^{DFM}\}$, is calculated using AIXRD in the same manner as for individual fragments previously. The total molecular scattering factor then becomes,

$$f_{\text{DFM}}(\mathbf{q}) = \left(\sum_{j=1}^{N_{\text{d}}} f_j^{DFM}(\mathbf{q}) \right) - f_{corr}, \tag{14}$$

where $N_{\text{d}} = N_{\text{f}}(N_{\text{f}} - 1)/2$ is the total number of possible dimers, and where f_{corr} is a correction factor that accounts for the double-counting inherent in the summation of the dimers in Eq. (14), as shown schematically in Fig. 1. This correction factor can be approximated as $f_{corr} = (N_{\text{f}} - 2) f_{\text{IFM}}(\mathbf{q})$, which approximately removes the double-counting of fragments when summing all individual dimers. In general, the number of dimer terms in Eq. (14) can be reduced by only including a subset of all dimers, e.g. only dimers formed from nearby or adjacent fragments, making sure to subtract double-counted fragments as appropriate.

It is appropriate to consider the equivalent approximations to the total electron density being made. Eqs. (13) and (14) corresponds precisely with total electron density approximations,

$$\rho_{\text{IFM}}(\mathbf{r}) = \sum_{j}^{N_{\text{f}}} \rho_j(\mathbf{r}), \tag{15}$$

and

$$\rho_{\text{DFM}}(\mathbf{r}) = \left(\sum_{k}^{N_{\text{f}}(N_{\text{f}}-1)/2} \rho_k(\mathbf{r}) \right) - (N_{\text{f}} - 2)\rho_{\text{IFM}}(\mathbf{r}), \tag{16}$$

where the subscripts j and k sum over fragments and dimers respectively. Similarly, the IAM electron density is simply,

$$\rho_{\mathrm{IAM}}(\mathbf{r}) = \sum_{i}^{N_{\mathrm{at}}} \rho_i(\mathbf{r}), \tag{17}$$

for N_{at} isolated atomic electron densities $\rho_i(\mathbf{r})$. It is perhaps worth emphasizing that the tabulated IAM scattering factors are based on isolated atoms (radicals), as validated in Fig. 1 in the *Supporting Information* which shows essentially exact agreement between tabulated IAM form factors and HF/6-31G* AIXRD calculations on isolated atoms. The IFM model currently proposed also uses radicals, however, as shown in the Results section below, radical fragments composed of multiple atoms constitute a significant improvement over single-atom radical fragments.

3 Results

We begin the Results section with an investigation of the influence of the selection of fragments in subsection 3.1, and then make a detailed comparison between the *ab initio* x-ray diffraction (AIXRD) calculations, the independent atom model (IAM), the independent (IFM) and the corrected dimer fragment (DFM) models in subsection 3.2. Throughout we use independent atom model form factors calculated using Hartree-Fock AIXRD calculations for reference (see *Supporting Information*). Finally, we investigate the scaling properties of the various methods in subsection 3.3. The elastic scattering (diffraction) is calculated for x-ray photons with energy 10.332 keV (1.2 Å) throughout this entire article.

3.1 Fragment selection in ethanol

To assess the effect of small fragments and examine the importance of fragment selection, the molecule ethanol was chosen.[49] Fragment selection in such a small molecule as ethanol is problematic, which makes it an excellent test-case. Two choices of fragments and dimers are tested to quantify the effect of assigning charges to the fragments. Ethanol is geometry optimized at the HF/6-311++G** level of theory, and this geometry is used throughout with the molecular scattering factor calculated using all-molecule AIXRD at the HF/6-31G* level used as reference.

The ethanol molecule is broken into three fragments and three dimers, as defined in Table 1 and visualized in Fig. 2. Two different fragment definitions are chosen: the first (IFM$^{\mathrm{a}}$) consists of the CH_3 end-group (fragment a), the CH_2 central group (fragment b), and the OH end group (fragment c); the second definition (IFM$^{\mathrm{b}}$) is similar except that the end groups are charged, i.e. fragment a' is CH_3^+, fragment c' is OH^-, and fragment b' is the same as fragment b. The IFM$^{\mathrm{a}}$ definition involves fragments a and c as doublet ground states because they have unpaired electrons, whereas IFM$^{\mathrm{b}}$ avoids this by introducing charged fragments a' and c'. The corresponding set of dimers, as defined in Table 1, are ab and ab', $ac = ac'$, and bc and bc', denoted DFM$^{\mathrm{a}}$ and DFM$^{\mathrm{b}}$ respectively. The Mulliken population analysis from the full HF/6-31G* calculation assigns charges of 0.017e, 0.282e, -0.299e to the CH_3, CH_2 and OH sections of the molecule. For this reason the neutral charged fragments seem a better choice at this stage. However, it is unclear which will perform better in the x-ray diffraction calculations.

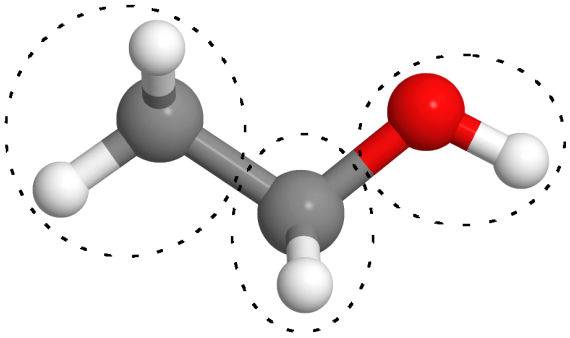

Figure 2: The molecule ethanol with the three fragments ($N_{\mathrm{f}} = 3$) shown. Fragments from left to right: a, b, and c. See also Table 1.

3.1.1 X-ray diffraction from ethanol

The ethanol molecular scattering factors for each method are calculated and compared to the reference *ab initio* scattering factor. The top figure in

Table 1: Two definitions of the three ethanol fragments ($N_f = 3$) and corresponding dimers for the IFM and DFM methods, showing labels, chemical formulae, the number of electrons, net charge, and spin multiplicity (see also Fig. 2).

Method	Label	Type	N_{el}	Q	$2S+1$
		Fragments			
IFM[a]	a	CH_3	9	0	2
	b	CH_2	8	0	3
	c	OH	9	0	2
IFM[b]	a'	CH_3^+	8	1	1
	$b' = b$	CH_2	8	0	3
	c'	OH^-	10	-1	1
		Dimers			
DFM[a]	ab	$a + b$	17	0	2
	ac	$a + c$	18	0	3
	bc	$b + c$	17	0	2
DFM[b]	ab'	$a' + b'$	16	1	1
	$ac' = ac$	$a' + c'$	18	0	3
	bc'	$b' + c'$	18	-1	1

Fig. 3 shows rotationally-averaged diffraction intensities for each method in the range $0 \leq q \leq 10$ Å^{-1}. The intensity scale on the y-axis is logarithmic to better visualize low intensity peaks. There are distinct shoulder peaks at $q = 1.7$, 3.1, 4.6, and 7.2 Å^{-1}, corresponding to real-space distances $r = 2\pi/q = 3.70$, 2.03, 1.37, and 0.87 Å respectively. These distances correspond to peaks in the gas-phase radial distribution function arising from the distances between atoms. The inset shows the minor shoulder peak at $q = 1.7$ Å^{-1} and closely reveals the differences between each method in this vicinity, with IAM furthest from the full AIXRD method, especially in the range $1.3 \leq q \leq 2.2$ Å^{-1}, implying the worst spatial representation of the electron density in the distance region of approximately 3.9 ± 1.0 Å.

To gain further insight about the differences between the methods, the top panel of the bottom figure in Fig. 3 shows the absolute difference signal, $|\Delta I(q)|$, defined as,

$$|\Delta I(q)| = |I_{\text{ref}}(q) - I_{\text{method}}(q)|, \tag{18}$$

where $I_{\text{ref}}(q)$ is the reference signal (normally the full AIXRD calculation) and $I_{\text{method}}(q)$ is the method being evaluated. The largest difference is at $q = 0.65$ Å^{-1} corresponding to $r = 9.7$ Å, although this difference only appears for IAM, IFM[a], and IFM[b], showing that particularly IAM, followed by IFM[b] then IFM[a], poorly represent longer range electronic distances. There is also a shoulder peak at ~ 1.6 Å^{-1} which is particularly significant for IAM, and to a lesser extent for DFM[b] and IFM[a].

It is helpful, especially at large values of q, to investigate the absolute percent difference signal, $|\%\Delta I(q)|$, defined as,

$$|\%\Delta I(q)| = 100 \times \frac{|\Delta I(q)|}{I_{\text{ref}}(q)}, \tag{19}$$

which is shown in the bottom panel of the bottom figure in Fig. 3. In percentage terms, there is a large difference peak at $q = 1.7$ Å^{-1}, the largest difference being IAM (9.91%), followed by IFM[a] (3.96%), then IFM[b], DFM[b], and finally DFM[a].

To complement Fig. 3, Table 2 shows mean and maximum values of $|\Delta I(q)|$ and $|\%\Delta I(q)|$ for each method. We assess the accuracy of the calculations using the mean and maximum difference values. The DFM[a] method has the lowest mean absolute and percentage error with only 0.43% mean error, it also has the lowest maximum error. The IAM method is the worst approximation, with quite significant maximum error of 9.91% at $q = 1.7$ Å^{-1} as previously mentioned. Overall, DFM performs better than IFM, and IFM and DFM both perform better than IAM by a substantial margin. The IFM[a] and IFM[b] methods perform similarly, but DFM[a] performs significantly better than DFM[b], showing that use of neutral charge fragments are a more appropriate choice for ethanol.

In addition to fully isotropic samples, molecules can be aligned to obtain extra information along the azimuthal scattering angle ϕ as well as along the radial angle θ. Fig. 4 shows the absolute difference signal, $|\Delta I(\mathbf{q}(\theta, \phi))|$ as defined by Eq. (18), for an aligned ethanol molecule with the full AIXRD calculation used as reference. Since the signal on the detector is anisotropic for an aligned molecule, the signal depends on the polar coordinates (θ, ϕ) rather than just the radial coordinate $q(\theta)$ with the relationship between q and θ given by Eq. (3). The incident x-ray wavevector direc-

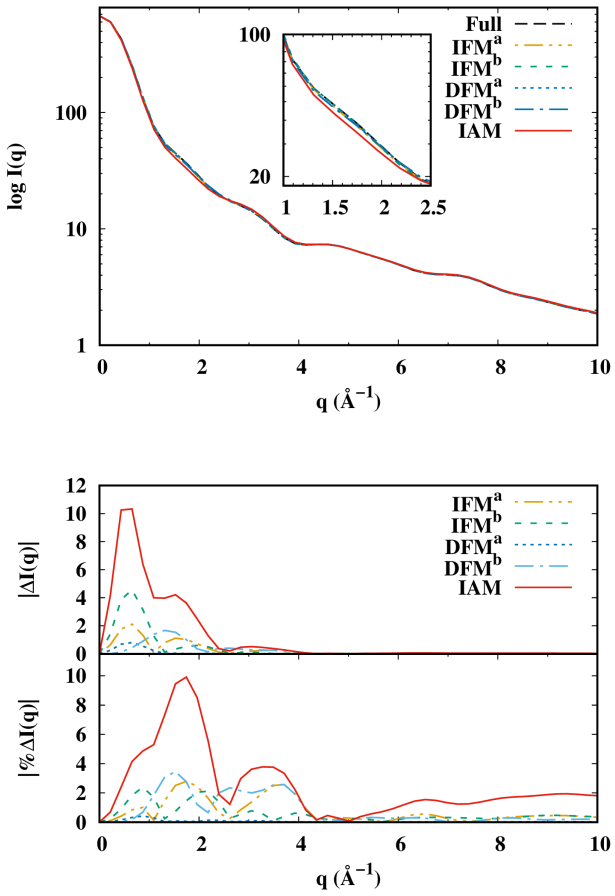

Figure 3: Rotationally-averaged diffraction from ethanol. (Top) Absolute diffraction signals shown on a log-scale for the intensity on the y-axis. The inset shows a close-up in the range $1.0 \leq q \leq 2.5$ Å^{-1}. The data for 'Full' corresponds to the full AIXRD reference calculation. (Bottom) Comparison of the various methods with the AIXRD reference calculation. The upper panel shows the absolute difference signals $|\Delta I(q)|$ calculated according to Eq. (18), and the bottom panel shows the same data represented as a percentage differences $|\%\Delta I(q)|$ according to Eq. (19).

tion is along the z-axis, which is perpendicular to the plane of the paper in Fig. 2. The centre of the circle corresponds to q=0 and the edge to q=10 Å^{-1}, with the corresponding radial scattering angle ranging from θ=0 to θ=π. Anti-clockwise around the circle is the azimuthal scattering angle $\phi = [0, 2\pi]$. It is clear that extra information is visible along the azimuthal angle for scattering from a perfectly aligned molecule, i.e. peaks in θ and ϕ arise from the positions of atoms (technically electron density) in 3D space, rather than

Table 2: Mean and maximum values of $|\Delta I(q)|$ (Eq. 18) and $|\%\Delta I(q)|$ (Eq. 19) for ethanol using the independent atom model and the fragment-based approximations with the full AIXRD calculations taken as reference.

Method	Mean		Max.	
	$\|\Delta I(q)\|$	$\|\%\Delta I(q)\|$	$\|\Delta I(q)\|$	$\|\%\Delta I(q)\|$
IFM^a	0.35	0.95	3.33	3.96
IFM^b	0.46	0.83	5.79	3.02
DFM^a	0.09	0.43	0.74	1.93
DFM^b	0.23	0.60	1.89	2.91
IAM	1.12	2.44	10.33	9.91

probabilistic radial distributions of the distances between atoms. The bottom figure shows the azimuthally integrated difference diffraction signals for each method, $\langle|\Delta I(q)|\rangle_\phi$, defined as,

$$\langle|\Delta I(q)|\rangle_\phi = \int_0^{2\pi} |\Delta I(\mathbf{q}(\theta, \phi))| \, \mathrm{d}\phi. \qquad (20)$$

The aligned results show that the DFM^a method is by far the closest to the reference results (by ~2 orders of magnitude) compared to the other methods. This emphasizes the importance of pair-wise fragment interaction corrections in small molecules, i.e. small fragments alone cannot significantly improve on the IAM approximation. As DFM^a is so much more accurate than DFM^b, it demonstrates that the definition of fragments a is a notably better representation of the electron density and thus the x-ray diffraction pattern. It also shows that the charge of each fragments is an important factor and can make or break the approximation depending on how the choice is made, noting further that DFM^a and DFM^b carry essentially the same computational cost.

The IAM method has the largest absolute error at q = 0.6 Å^{-1}, however has similar error to IFM^b for $q > 2$ Å^{-1}. Suprisingly it has lower error than IFM^a and DFM^b for $q > 4$ Å^{-1}. This shows that in some cases specific choices of fragments can end up worse than simply using the IAM method

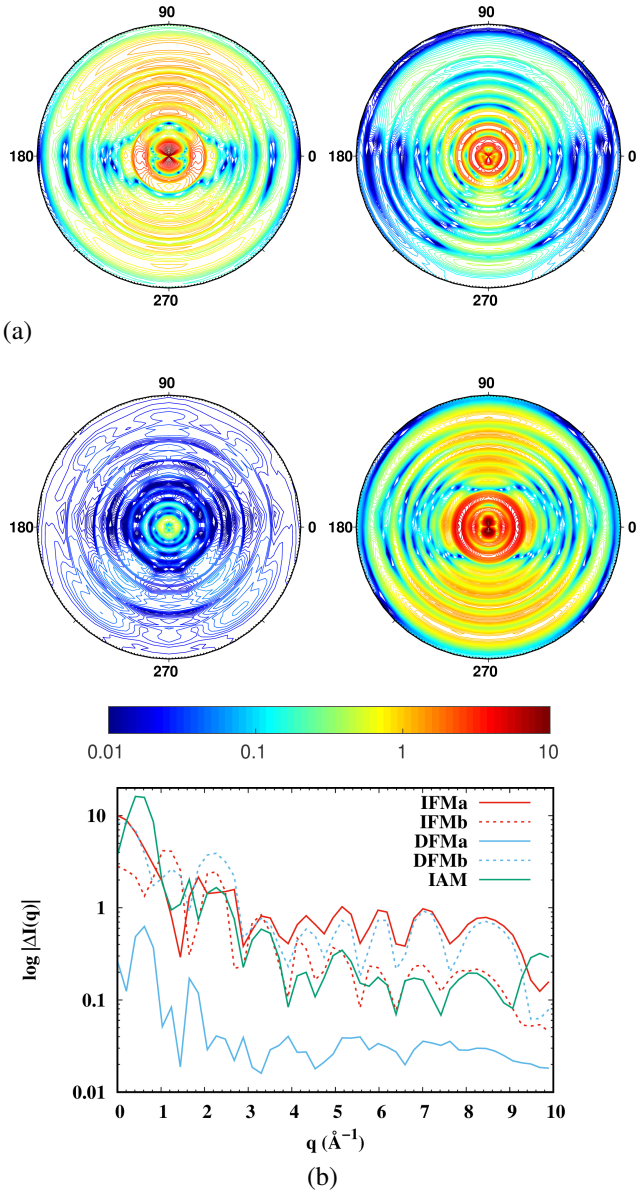

Figure 4: Diffraction from aligned ethanol molecules. (a) Absolute difference signals compared to the full AIXRD method, $|\Delta I(\mathbf{q}(\theta,\phi))| = |I_{\text{full}} - I_{\text{method}}|$ (Eq. 18). The centre of each circle is $q = 0$ and the outer edge of the circle is $q = 10$ Å^{-1}. Anti-clockwise around the the circle is the azimuthal angle ϕ. The colours represent a logscale, with blue as 10^{-2} and red as 10. Top left: IFM$^{\text{a}}$, top right: IFM$^{\text{b}}$, bottom left: DFM$^{\text{a}}$, bottom right: DFM$^{\text{b}}$. (b) The corresponding azimuthally integrated absolute difference signals, $\langle|\Delta I(q)|\rangle$ (Eq. 20). The lower the values, the better agreement with the reference.

at larger q values. Depending on the experiment however, the range $q < 4$Å^{-1} corresponds to a region of relatively high signal-to-noise, and noting that signal drops off exponentially with q, it is reasonable to choose to improve the accuracy of theoretical calculations in this range (low-q) even at the cost of some accuracy at high-q. Additionally, IFM$^{\text{a}}$ and DFM$^{\text{b}}$ are very similar in error, showing that including dimers does not always improve the situation if the charges are not assigned well. However, these types of problems can be expected to become less important as the size of the fragments increases. In that sense, the above example demonstrates that it is not really appropriate to divide a molecule as small as ethanol into fragments.

3.2 Fragment selection in diphenylalanine

Diphenylalanine (FF) is a peptide consisting of two phenylalanine residues. This peptide is used to assess the accuracy and efficiency of the fragment-based methods in organic biomolecules such as peptides, proteins, and DNA or RNA. Table 3 shows the definition of the fragments and dimers chosen for diphenylalanine, and Fig. 5 shows a schematic of the fragments. Two separate choices are made, one with $N_{\text{f}} = 4$ fragments (IFM^4), a, b, c, and d, and their $N_{\text{f}}(N_{\text{f}}-1)/2 = 6$ dimers, and the other definition with $N_{\text{f}} = 2$ fragments (IFM^2). The $N_{\text{f}} = 2$ method uses the combined fragments $a+b$ and $c+d$ and the only broken bond is the peptide (C-N) bond between the two phenylalanine groups. In this way the effect of fragment size can be systematically studied. The computations are compared to reference AIXRD calculations at the HF/6-31G* level.

3.2.1 X-ray diffraction from diphenylalanine

The top figure in Fig. 7 shows rotationally-averaged diffraction intensities on a log-scale for each method in the range $0 \leq q \leq 10$ Å^{-1}. The inset shows the range $1.0 \leq q \leq 3.5$ Å^{-1} and reveals the differences between the methods in this region, with IAM furthest removed from the full AIXRD reference results, implying that IAM has the worst spatial representation of the electron density in the real-space distance region of approximately $r = 2\pi/q = [1.8, 6.3]$ Å. The bottom part of Fig. 7 shows the same data but in

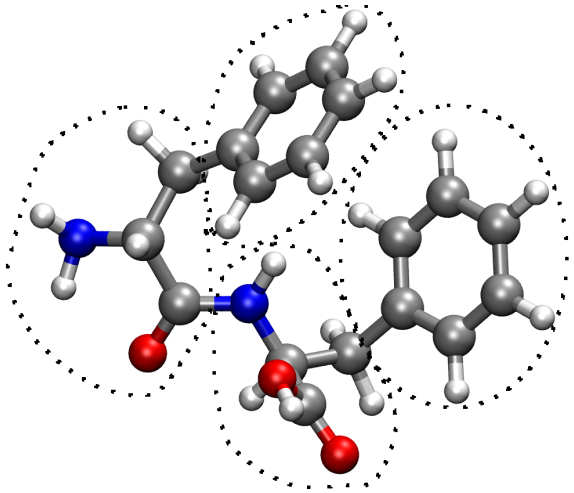

Figure 5: Diphenylalanine with four ($N_\mathrm{f} = 4$) defined fragments. The top phenyl group is labelled as a, with connecting CH_2CHNH_2CO as b, $NHCHCOOHCH_2$ as c, and the second phenyl group as d. The two-fragment ($N_\mathrm{f} = 2$) scenario simply corresponds to the two phenylalanine units (F+F). See also Table 3.

Table 3: Definition of diphenylalanine fragments and dimers, showing labels, chemical formula, number of atoms, number of electrons, charge, and spin multiplicity. Two cases are considered, four ($N_\mathrm{f} = 4$) and two ($N_\mathrm{f} = 2$) fragments, with DFM only relevant in the $N_\mathrm{f} = 4$ scenario. The net charge is zero for all fragments. See also Fig. 5.

Label	Type	N_atom	N_el	$2S+1$
IFM^4 ($N_\mathrm{f} = 4$)				
a	Phenyl·	11	41	2
b	$CH_2CHNH_2\dot{C}O$	10	38	1
c	$\dot{N}HCHCOOHCH_2$	11	46	1
d	Phenyl·	11	41	2
IFM^2 ($N_\mathrm{f} = 2$)				
ab	$a+b$	21	79	2
cd	$c+d$	22	87	2
DFM^4 ($N_\mathrm{f} = 4$)				
ab	$a+b$	21	79	2
ac	$a+c$	22	87	2
ad	$a+d$	22	82	1
bc	$b+c$	21	84	1
bd	$b+d$	21	79	2
cd	$c+d$	22	87	2

terms of absolute $|\Delta I(q)|$ (upper panel) and absolute percent differences $|\%\Delta I(q)|$ (lower panel). In the range $q = [1.0, 2.5]$ Å^{-1} IAM underestimates the diffraction intensity compared to the AIXRD method; in this region is the maximum percentage difference is $|\%\Delta I(q)| = 10.9\%$ at $q = 1.74$ Å^{-1} ($r = 3.6$ Å). Conversely, IAM overestimates $I(q)$ in the region $q = [2.5, 4.0]$ Å^{-1}; a distinct peak is here $|\%\Delta I(q)| = 5.8\%$ at $q = 3.05$ Å^{-1} ($r = 2.1$ Å). In terms of absolute difference, the largest $|\Delta I(q)|$ peak is at 0.22 Å^{-1} but this represents a low percentage difference of $|\%\Delta I(q)| = 0.42\%$, the second and third largest peaks are at 0.66 Å^{-1} and represents a larger $|\%\Delta I(q)|$ of 3.1%, and at 1.53 Å^{-1} with $|\%\Delta I(q)| = 9.5\%$ respectively (this overlaps with the largest $|\%\Delta I(q)| = 10.9\%$ peak). A final smaller $|\Delta I(q)|$ peak is at 3.05 Å^{-1}, overlapping with the second largest percent difference peak, $|\%\Delta I(q)| = 5.8\%$. Overall, this shows that the IAM method underestimates the electron density for distances in the range $[2.5, 6.3]$ Å and overestimates the density in the range $[1.6, 2.5]$ Å. In other words, the IAM representation of electron density is too sparse at longer distances and too dense at shorter distances. It also slightly overestimates the intensity going out from $q > 4.8$ Å^{-1} ($r < 1.3$ Å), corresponding to an increasing percentage difference of $|\%\Delta I(q)| = 1.0 - 2.2\%$ up to 10 Å^{-1}, this shows further that at shorter distances (close to the nuclei) IAM overestimates the electron density. This is understandable as within the IAM no redistribution of density is included and therefore more density resides close to the nuclei.

To illustrate further where the IAM misrepresents the electron density, Fig. 6 shows a close up of one of the phenyl rings in diphenylalanine with isosurfaces representing absolute electron density difference between HF/6-31G* and the IAM, $|\Delta\rho(\mathbf{r})| = |\rho_\mathrm{IAM}(\mathbf{r}) - \rho_\mathrm{HF}(\mathbf{r})|$. It reveals the radial anisotropy around the centre of each carbon atom arising from the quantum chemistry calculation accounting for the π-bonding system in the electron density. On the other hand, the IAM incorrectly assumes that the electron density is spherically symmetric around each atom, thus, does not account for π-bonding in phenyl-groups (as shown), lone pairs, double bonds, or indeed any chemical bonding which redistributes the elec-

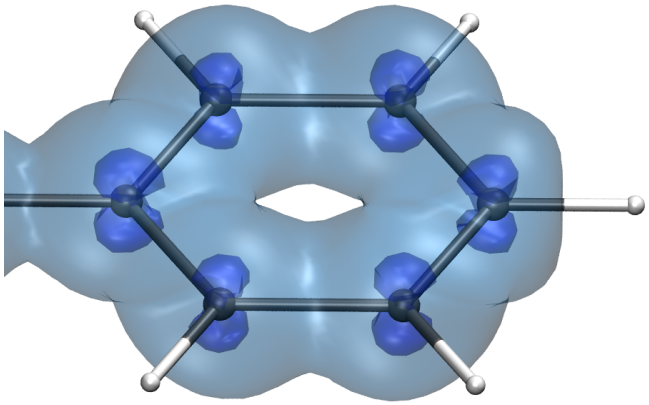

Figure 6: Close-up of one of the phenyl groups in diphenylalanine, showing the absolute difference between the HF/6-31G* and the IAM electron density, $|\Delta\rho(\mathbf{r})| = |\rho_{\mathrm{IAM}}(\mathbf{r}) - \rho_{\mathrm{HF}}(\mathbf{r})|$. Dark blue represents isosurfaces with $|\Delta\rho(\mathbf{r})| = 0.2$ and light blue represents $|\Delta\rho(\mathbf{r})| = 0.1$.

trons away from isotropic shells around each individual atom.

Table 4: Mean and maximum values of $|\Delta I(q)|$ (Eq. 18) and $|\%\Delta I(q)|$ (Eq. 19) for diphenylalanine using the independent atom model (IAM) and the fragment-based approximations (IFM^2, IFM^2, DFM^4) with the full AIXRD calculations taken as reference.

Method	Mean		Max	
	$\lvert\Delta I(q)\rvert$	$\lvert\%\Delta I(q)\rvert$	$\lvert\Delta I(q)\rvert$	$\lvert\%\Delta I(q)\rvert$
IFM^2	0.35	0.11	6.26	0.52
IFM^4	0.70	0.23	11.84	1.23
DFM^4	0.22	0.07	3.47	0.51
IAM	6.49	2.62	59.56	10.92

Table 4 shows mean and maximum values of $|\Delta I(q)|$ and $|\%\Delta I(q)|$ for each method. The DFM^4 method has the lowest mean absolute and percentage error with only 0.07% mean error, it also has the lowest maximum errors. As before, the IAM method is the worst approximation, with a large maximum error of 10.92% at $q = 1.9$ Å^{-1}. This is similar to the IAM difference in the ethanol results in terms of mean error and position in q of errors. The IFM^2 method has about half the absolute and percentage errors compared to IFM^4 showing the benefit of using larger fragments. It also performs reasonably close to the DFM^4 method with a

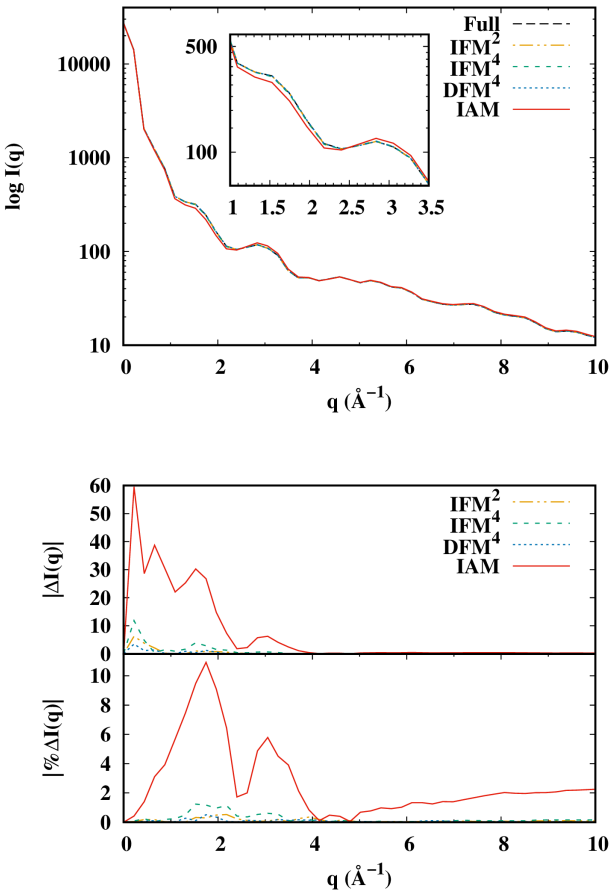

Figure 7: Rotationally-averaged diffraction from diphenylalanine. (Top) Absolute diffraction signals shown on a log-scale for the intensity on the y-axis. The inset shows a close-up in the range $1.0 \leq q \leq 3.5$ Å^{-1}. The data for 'Full' corresponds to the full AIXRD reference calculation. (Bottom) Comparison of the various methods with the AIXRD reference calculation. The upper panel shows the absolute difference signals $|\Delta I(q)|$ calculated according to Eq. (18), and the bottom panel shows the same data represented as a percentage differences $|\%\Delta I(q)|$ according to Eq. (19).

mean percentage error of 0.11%. Noting that DFM costs far more than IFM, it may be worthwhile using IFM with fragment sizes consisting of about 20 atoms, especially for larger molecules.

Fig. 8 shows the absolute difference signal $|\Delta I(\mathbf{q}(\theta, \phi))|$ (Eq. 18) for aligned diphenylalanine. The incident x-ray wavevector direction is along the z-axis, which is perpendicular to the plane of the paper in Fig. 5. As before, from the centre of the circle to the edge is $q \in [0, 10]$ Å^{-1}, corresponding to radial scattering angle range $\theta =$

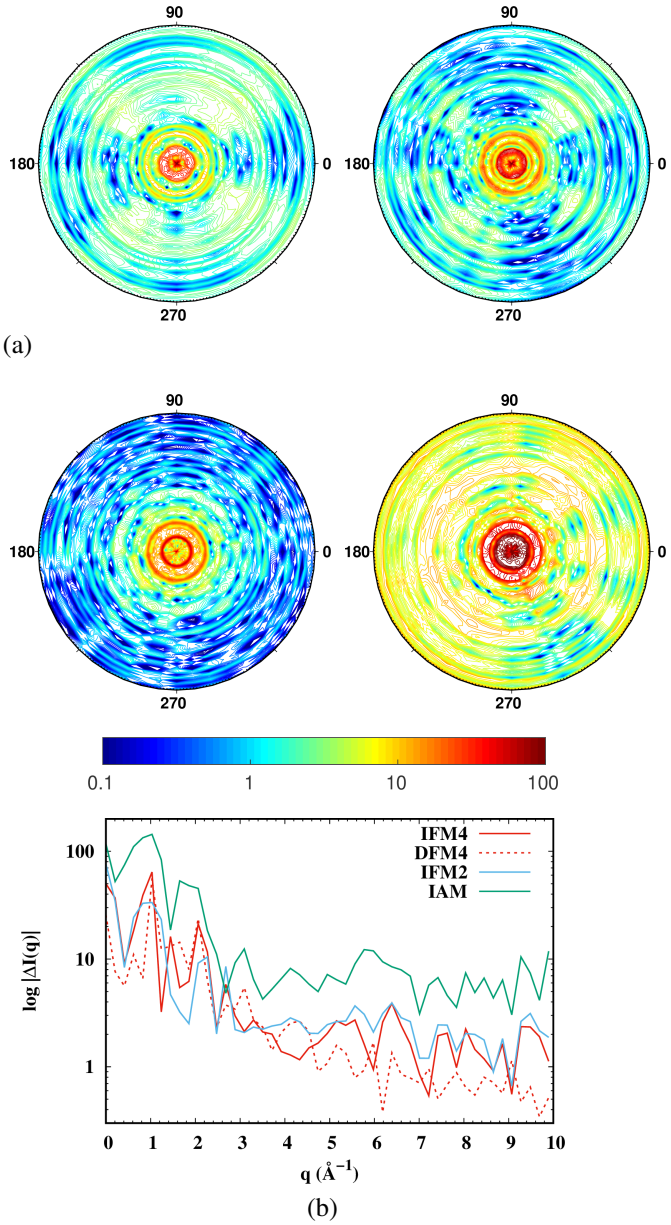

Figure 8: Diffraction from aligned diphenylalanine molecules. (a) Absolute difference signals compared to the full AIXRD method, $|\Delta I(\mathbf{q}(\theta,\phi))| = |I_{\text{full}} - I_{\text{method}}|$ (Eq. 18). The centre of each circle is $q = 0$ and the outer edge of the circle is $q = 10$ Å^{-1}. Anti-clockwise around the the circle is the azimuthal angle ϕ. The colours represent a logscale, with blue as 0.1 and red as 100. Top left: IFM^2, top right: IFM^4, bottom left: DFM^4, bottom right: IAM. (b) The corresponding azimuthally integrated absolute difference signals, $\langle|\Delta I(q)|\rangle$ (Eq. 20). The lower the values, the better agreement with the reference.

$[0,\pi]$. Anti-clockwise around the circle is the azimuthal scattering angle $\phi = [0, 2\pi]$.

As expected from the rotationally-averaged results, it is apparent that the IAM is furthest from the reference by about one order of magnitude compared to the other methods, depending on q. The DFM^4 method is best overall (at all values of q), however it is very similar to IFM^4 and IFM^2 at $q < 4.5$ Å^{-1}, and it is only somewhat better at $q > 4.5$ Å^{-1}. Additionally, IFM^2 and IFM^4 perform similarly with IFM^2 performing better at $q \sim 1 - 2$Å^{-1} and IFM^4 slightly better at $q > 3.5$Å^{-1}. Taking into consideration the computational cost of DFM, it may be worth generally employing IFM^2 in this case as it performs similarly overall at a fraction of the cost, as will be discussed further in the next section.

3.3 Scaling of computations

It is relevant to quantify the computational savings possible with the IFM and DFM methods compared to a full *ab initio* method such as AIXRD. We examine the scaling of the equations first, and then the actual timing data for the diphenylalanine calculations. The time required for a IFM calculation is,

$$t_{\text{IFM}} = \sum_{j}^{N_{\text{f}}} t_j, \tag{21}$$

where t_j is the AIXRD timing for the jth fragment. The linearization of the computations reflected by Eq. (21) is achieved by the decomposition of the molecule into fragments and underpins the computational efficiency of IFM, in a similar manner to IAM. The scaling of the DFM method is given by,

$$t_{\text{DFM}} = t_{\text{IFM}} + \sum_{k}^{N_{\text{d}}} t_k, \tag{22}$$

where t_k is the AIXRD timing for the kth dimer. It is useful to note that dimers tend to be twice the size of the fragments, and because of the non-linear scaling of AIXRD discussed below, $t_k \sim (2t_j)^h$, with $1 < h \leq 2$.

In AIXRD, the total time taken is,

$$t_{\text{AIXRD}} \propto N_{\text{MO}}(N_{\text{BF}}N_{\text{g}})^h, \tag{23}$$

for N_{MO} occupied molecular orbitals, N_{BF} basis functions (or contractions) per MO, and N_{g} GTOs per basis function. In terms of the number of GTOs per atom, we get,

$$t_{\mathrm{AIXRD}} \propto N_{\mathrm{MO}}(N_{\mathrm{C}}N_{\mathrm{GTO/C}} + N_{\mathrm{H}}N_{\mathrm{GTO/H}})^h, \quad (24)$$

for N_{MO} orbitals, N_{C} carbon-like atoms (i.e. B, C, N, O, F), and N_{H} hydrogens in the molecule (or fragment), with $N_{\mathrm{GTO/C}}$ and $N_{\mathrm{GTO/H}}$ GTOs per carbon-like atom and per hydrogen atom respectively; both of which are defined by the basis set. For example, in the 6-31G* basis set, each carbon-like atom has 6 + 3 + 1 *s*-orbitals, 3 + 1 *p*-orbitals (for p_x, p_y, p_z), and 1 *d*-orbital (for the 6 Cartesian *d*-shells), and each hydrogen has 3 + 1 *s*-orbitals, therefore, $N_{\mathrm{GTO/C}} = 10 + 3(4) + 6(1) = 28$ and $N_{\mathrm{GTO/H}} = 4$. Note that,

$$N_{\mathrm{BF}}N_{\mathrm{g}} = N_{\mathrm{C}}N_{\mathrm{GTO/C}} + N_{\mathrm{H}}N_{\mathrm{GTO/H}} = N_{\mathrm{GTO}}, \quad (25)$$

i.e. the number of GTOs per MO, N_{GTO}, equals the number of carbon-like GTOs plus the number of hydrogen GTOs. This is useful to consider because a fragment has a lower number of atoms than the entire molecule, and the scaling depends non-linearly on the terms $N_{\mathrm{GTO/C}}$ and $N_{\mathrm{GTO/H}}$, thus, fragmentation reduces the scaling substantially.

In practice, a cut-off C_{cutoff} is used to skip calculations involving Gaussian products which have small overlap, i.e. when the basis set coefficients c_i, c_j or orbital coefficients M_i, Mj cause the Gaussian product term in the electron density (Eq. 5) to be less than C_{cutoff}. This reduces the total number of Gaussian products that must be computed, and lowers the number of terms from $(N_{\mathrm{BF}}N_{\mathrm{g}})^2$ to $(N_{\mathrm{BF}}N_{\mathrm{g}})^h$, with $1<h\leq 2$. The effective scaling becomes $h\sim 1.8$ when a cut-off value of $C_{\mathrm{cutoff}} = 10^{-9}$ is used, which effectively retains the accuracy of full AIXRD. Increasing C_{cutoff} further would lower h, but at the cost of the overall accuracy of the calculations.

Table 5 shows time taken for the diphenylalanine calculations using full AIXRD, IFM^2, IFM^4, and DFM^4. An Intel Xeon E5-2620 (2.10GHz) CPU was used in parallel on 8 cores. It is clear that the IFM^2 method does not reduce computational effort in the present example. However, the IFM^4 calculation is a factor 3.4 times faster than full AIXRD and, as shown in Section 3.2, it provides a significant improvement on IAM. Finally, the DFM^4 method is not worth performing in this case as it is much slower than full AIXRD.

Table 5: Timings for diphenylalanine calculations on an Intel Xeon E5-2620 processor.

Method	Time (hrs)
Full AIXRD	9.4
IFM^2	8.5
IFM^4	2.8
DFM^4	21.3

It is important to note that the computational saving for IFM becomes greater for larger molecules with a greater number of fragments. For example, using IFM instead of AIXRD on a protein with 1000 carbon atoms (ignoring hydrogens for simplicity) split into $N_{\mathrm{f}} = 100$ fragments, would scale as $t_{\mathrm{IFM}} \propto N_{\mathrm{f}}(10)^h$ with 10 carbons per fragment on average (see Eq. 24), whereas full AIXRD would scale as $t_{\mathrm{Full}} \propto 1000^h$ giving $t_{\mathrm{Full}}/t_{\mathrm{IFM}} = N_{\mathrm{f}}^{h-1}$, which clearly favours IFM for large values of N_{f}.

Lastly, we note that it is normally sufficient to consider substantially fewer than $N_{\mathrm{d}} = N_{\mathrm{f}}(N_{\mathrm{f}} - 1)/2$ dimers. This is because it is generally a very good approximation to define dimers only for adjacent fragments. In this case, the DFM method is cheaper than full AIXRD, but even in the ideal case of a long repeating unit molecule such as a polymer chain, the number of dimers would be approximately the same as the number of fragments. Noting that each dimer costs $\approx (2t_{\mathrm{f}})^h$, where t_{f} is the average timing of a fragment, the expense is still very high. In terms of cost effectiveness, the most beneficial strategy is therefore to use IFM with sufficiently large fragments to avoid any significant loss of accuracy compared to DFM.

4 Conclusions

The presented fragment method is capable of reproducing *ab initio* x-ray diffraction data to a high accuracy and scales well to large molecules. The method could also be adapted for electron diffraction.[20,34,57] Our results emphasize the shortcom-

ings of the independent atom model (IAM), with significant improvements resulting from the presented independent fragment model (IFM). The pair-wise corrections introduced by the dimer fragment model (DFM) are useful, but are most valuable when the selected fragments are small. A computationally sensible strategy is therefore to use IFM without pair-wise corrections, but instead to select as large fragments as possible, which minimizes the issues originating in small fragments demonstrated. One could also consider using e.g. DFT instead of HF to support even larger fragments. It is important to note that the included reference calculations provide us with an exact understanding of the magnitude of the shortcomings of the currently implemented IFM approach.

Furthermore, we emphasize that the IFM approach carries significant computational advantages, in part because it linearizes the problem, but also because fragment form factors can be precalculated and tabulated, allowing accurate all-molecule form factors to be constructed with great computational efficiency in an almost lego-like manner. It is exactly this computational advantage that has made the independent atom model so successful, despite some of its obvious shortcomings. The approach presented here thus replicates the greatest advantages of IAM, while also greatly improving upon the quality of the predicted scattering.

The computational efficiency gains achieved by introducing a cut-off into the AIXRD calculations are striking and essentially mimick the fragment-based approach in an adaptive manner. Pragmatically, the cut-off might therefore be the most useful approach in many situations. A fragment-based strategy, such as IFM, is likely to be most useful in very large molecules such as polymers with a high degree of repeating (and ideally rigid) subunits where a library of pre-calculated fragment form factors can be exploited maximally. Obvious examples includes peptides, proteins, and DNA or RNA. We also highlight the close conceptual relationship between our IFM method and the MEDLA approach for peptide electron densities[58] and the highly-coarse-grained MARTINI-beads approach for the prediction of small-angle x-ray scattering solution-phase signals.[59]

An alternative to the presented hybrid IFM-AIXRD approach is to try to obtain full-molecule wavefunctions from *ab initio* electronic structure methods capable of treating very large molecules. The already discussed FMO[48,49,60] method is a strong contender and should be explored, but DFT-based calculations are also an option, although convergence issues appear pervasive when dealing with DFT for large polypeptides[61–63] and it would be important to use long-range corrected functionals with exchange. The decreasing computational costs associated with highly parallel GPU-based electronic structure codes may also provide an alternative.[61] In previous AIXRD studies we highlighted the energy convergence of the *ab initio* calculations as a proxy for convergence in the scattering calculations.[37] In the current fragment-based approach, the total energy is a less useful predictor of net convergence in the scattering, which is a further argument in favour of full-molecule calculations using e.g. FMO or DFT.

In terms of an outlook for the presented IFM-AIXRD approach, an obvious continuation is to improve the level of electronic structure theory beyond Hartree-Fock (HF). We have used HF in this study as it matches the accuracy of tabulated atomic form factors,[54–56] but the level of electronic structure theory is known to affect e.g. the stable conformations of peptides[64] and the absolute level of x-ray scattering convergence.[37] Furthermore, multiconfigurational methods would allow for accurate predictions of total scattering.[35] In the context of dynamics, our approach is likely to be most valuable for dynamics of entire macromolecules, for instance as in the study of protein quakes in Ref.[12] or in studies of chromophores embedded in a matrix such as a protein. We anticipate that the greatest impact might be in the context of interpreting single-molecule x-ray scattering data for structure determination of biomolecules, utilizing *ab initio* and experimental data for refinement.[65]

Acknowledgments

A.K. acknowledges support from a Royal Society of Edinburgh Sabbatical Fellowship (58507) and a research grant from the Carnegie Trust for the Universities of Scotland (CRG050414). The authors thank Dmitri G. Fedorov for helpful discussions

regarding FMO theory.

Supporting Information

A comparison between AIXRD calculations and atomic form factors, as well as published experimental gas-phase x-ray scattering data for 1,3-cyclohexadiene, is given. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Emma, P.; Akre, R.; Arthur, J.; Bionta, R.; Bostedt, C.; Bozek, J.; Brachmann, A.; Bucksbaum, P.; Coffee, R.; Decker, F.-J. et al. First lasing and operation of an Ångström-wavelength free-electron laser. Nat. Phot. **2010**, 4, 641–647.

(2) Ackermann, W.; Asova, G.; Ayvazyan, V.; Azima, A.; Baboi, N.; Bähr, J.; Balandin, V.; Beutner, B.; Brandt, A.; Bolzmann, A. et al. Operation of a free-electron laser from the extreme ultraviolet to the water window. Nat. Phot. **2007**, 1, 336–342.

(3) Ishikawa, T.; Aoyagi, H.; Asaka, T.; Asano, Y.; Azumi, N.; Bizen, T.; Ego, H.; Fukami, K.; Fukui, T.; Furukawa, Y. et al. A compact X-ray free-electron laser emitting in the sub-Ångström region. Nat. Phot. **2012**, 6, 540–544.

(4) Allaria, E.; Appio, R.; Badano, L.; Barletta, W.; Bassanese, S.; Biedron, S.; Borga, A.; Busetto, E.; Castronovo, D.; Cinquegrana, P. et al. Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet. Nat. Phot. **2012**, 6, 699–704.

(5) Johansson, L. C.; Stauch, B.; Ishchenko, A.; Cherezov, V. A bright future for serial femtosecond crystallography with XFELs. Trends Biochem. Sci. **2017**, 42, 749–762.

(6) Minitti, M. P.; Budarz, J. M.; Kirrander, A.; Robinson, J. S.; Ratner, D.; Lane, T. J.; Zhu, D.; Glownia, J. M.; Kozina, M.; Lemke, H. T. et al. Imaging Molecular Motion: Femtosecond X-Ray Scattering of an Electrocyclic Chemical Reaction. Phys. Rev. Lett. **2015**, 114, 255501.

(7) Budarz, J. M.; Minitti, M. P.; Cofer-Shabica, D. V.; Stankus, B.; Kirrander, A.; Hastings, J. B.; Weber, P. M. Observation of femtosecond molecular dynamics via pump-probe gas phase X-ray scattering. J. Phys. B **2016**, 49, 034001.

(8) Glownia, J. M.; Natan, A.; Cryan, J. P.; Hartsock, R.; Kozina, M.; Minitti, M. P.; Nelson, S.; Robinson, J.; Sato, T.; van Driel, T. et al. Self-Referenced Coherent Diffraction X-Ray Movie of Ångström- and Femtosecond-Scale Atomic Motion. Phys. Rev. Lett. **2016**, 117, 153003.

(9) Yong, H.; Zotev, N.; Stankus, B.; Ruddock, J. M.; Bellshaw, D.; Boutet, S.; Lane, T. J.; Liang, M.; Carbajo, S.; Robinson, J. S. et al. Determining Orientations of Optical Transition Dipole Moments Using Ultrafast X-ray Scattering. J. Phys. Chem. Lett. **2018**, 9, 6556–6562.

(10) Ruddock, J. M.; Zotev, N.; Stankus, B.; Yong, H.-W.; Bellshaw, D.; Boutet, S.; Lane, T. J.; Liang, M.; Carbajo, S.; Du, W. et al. Simplicity beneath Complexity: Counting Molecular ElectronsReveals Transients and Kinetics of Photodissociation Reactions. Angew. Chem. Int. Ed. **2019**, 0, null.

(11) Levantino, M.; Schirò, G.; Lemke, H. T.; Cottone, G.; Glownia, J. M.; Zhu, D.; Chollet, M.; Ihee, H.; Cupane, A.; Cammarata, M. Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. Nat. Comm. **2015**, 6.

(12) Arnlund, D.; Johansson, L. C.; Wickstrand, C.; Barty, A.; Williams, G. J.; Malmerberg, E.; Davidsson, J.; Milathianaki, D.; DePonte, D. P.; Shoeman, R. L. et al. Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. Nat. Meth. **2014**, 11, 923–926.

(13) Chollet, M.; Alonso-Mori, R.; Cammarata, M.; Damiani, D.; Defever, J.; Delor, J. T.; Feng, Y.; Glownia, J. M.; Langton, J. B.; Nelson, S. et al. The X-ray pump–probe instrument at the linac coherent light source. J. Synchr. Rad. **2015**, 22, 503–507.

(14) Pemberton, C. C.; Zhang, Y.; Saita, K.; Kirrander, A.; Weber, P. M. From the (1B) spectroscopic state to the photochemical product of the ultrafast ring-opening of 1,3-cyclohexadiene: a spectral observation of the complete reaction path. J. Phys. Chem. A **2015**, 119, 8832–8845.

(15) Tudorovskaya, M.; Minns, R. S.; Kirrander, A. Effects of probe energy and competing pathways on time-resolved photoelectron spectroscopy: the ring-opening of 1,3-cyclohexadiene. Phys. Chem. Chem. Phys. **2018**, 20, 17714–17726.

(16) Smith, A. D.; Warne, E. M.; Bellshaw, D.; Horke, D. A.; Tudorovskya, M.; Springate, E.; Jones, A. J. H.; Cacho, C.; Chapman, R. T.; Kirrander, A. et al. Mapping the Complete Reaction Path of a Complex Photochemical Reaction. Phys. Rev. Lett. **2018**, 120, 183003.

(17) Stankus, B.; Zotev, N.; Rogers, D. M.; Gao, Y.; Odate, A.; Kirrander, A.; Weber, P. M. Ultrafast photodissociation dynamics of 1,4-diiodobenzene. J. Chem. Phys. **2018**, 148, 194306.

(18) Ischenko, A. A.; Weber, P. M.; Miller, R. J. D. Capturing Chemistry in Action with Electrons: Realization of Atomically Resolved Reaction Dynamics. Chem. Rev. **2017**, 117, 11066–11124.

(19) Centurion, M. Ultrafast imaging of isolated molecules with electron diffraction. J. Phys. B **2016**, 49, 062002.

(20) Stefanou, M.; Saita, K.; Shalashilin, D. V.; Kirrander, A. Comparison of Ultrafast Electron and X-Ray Diffraction A Computational Study. Chem. Phys. Lett. **2017**, 683, 300–305.

(21) Wolf, T. J. A.; Sanchez, D. M.; Yang, J.; Parrish, R. M.; Nunes, J. P. F.; Centurion, M.; Coffee, R.; Cryan, J. P.; Gühr, M.; Hegazy, K. et al. Imaging the Photochemical Ring-Opening of 1,3-Cyclohexadiene by Ultrafast Electron Diffraction. Nat. Chem. **2019**,

(22) Seibert, M. M.; Ekeberg, T.; Maia, F. R.; Svenda, M.; Andreasson, J.; Jönsson, O.; Odić, D.; Iwan, B.; Rocker, A.; Westphal, D. et al. Single mimivirus particles intercepted and imaged with an X-ray laser. Nature **2011**, 470, 78–81.

(23) Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U. et al. Femtosecond X-ray protein nanocrystallography. Nature **2011**, 470, 73–77.

(24) Redecke, L.; Nass, K.; DePonte, D. P.; White, T. A.; Rehders, D.; Barty, A.; Stellato, F.; Liang, M.; Barends, T. R. M.; Boutet, S. et al. Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser. Science **2013**, 339, 227–230.

(25) Barends, T. R. M.; Foucar, L.; Botha, S.; Doak, R. B.; Shoeman, R. L.; Nass, K.; Koglin, J. E.; Williams, G. J.; Boutet, S.; Messerschmidt, M. et al. De novo protein crystal structure determination from X-ray free-electron laser data. Nature **2014**, 505, 244–247.

(26) Helliwell, J. R. How to solve protein structures with an X-ray laser. Science **2013**, 339, 146–147.

(27) Garman, E. F. Developments in X-ray crystallographic structure determination of biological macromolecules. Science **2014**, 343, 1102–1108.

(28) Gati, C.; Oberthuer, D.; Yefanov, O.; Bunker, R. D.; Stellato, F.; Chiu, E.; Yeh, S.-M.; Aquila, A.; Basu, S.; Bean, R. et al. Atomic structure of granulin determined

from native nanocrystalline granulovirus using an X-ray free-electron laser. Proc. Nat. Acad. Sci. **2017**, 114, 2247–2252.

(29) von Ardenne, B.; Mechelke, M.; Grubmüller, H. Structure determination from single molecule X-ray scattering with three photons per image. Nat. Comm. **2018**, 9, 2375.

(30) Gatti, C., Macchi, P., Eds. Modern Charge-Density Analysis, 1st ed.; Springer Netherlands, 2012.

(31) Als-Nielsen, J.; McMorrow, D. Elements of Modern X-ray Physics; John Wiley & Sons, 2011.

(32) Küpper, J.; Stern, S.; Holmegaard, L.; Filsinger, F.; Rouzeée, A.; Rudenko, A.; Johnsson, P.; Martin, A. V.; Adolph, M.; Aquila, A. et al. X-Ray Diffraction from Isolated and Strongly Aligned Gas-Phase Molecules with a Free-Electron Laser. Physical Review Letters **2014**, 112, 083002.

(33) Northey, T.; Zotev, N.; Kirrander, A. Ab initio calculation of molecular diffraction. J. Chem. Theory Comp. **2014**, 10, 4911–4920.

(34) Carrascosa, A. M.; Kirrander, A. *Ab initio* calculation of inelastic scattering. Phys. Chem. Chem. Phys. **2017**,

(35) Carrascosa, A. M.; Yong, H.; Crittenden, D. L.; Weber, P. M.; Kirrander, A. Ab-initio calculation of total x-ray scattering from molecules. J. Chem. Theory Comp. **2019**, 0, null.

(36) Kirrander, A. X-ray diffraction assisted spectroscopy of Rydberg states. J. Chem. Phys. **2012**, 137, 154310.

(37) Northey, T.; Moreno Carrascosa, A.; Schäfer, S.; Kirrander, A. Elastic X-ray scattering from state-selected molecules. J. Chem. Phys. **2016**, 145, 154304.

(38) Northey, T. Ab Initio Molecular Diffraction; The University of Edinburgh, 2017.

(39) Carrascosa, A. M.; Northey, T.; Kirrander, A. Imaging rotations and vibrations in polyatomic molecules with X-ray scattering. Phys. Chem. Chem. Phys. **2017**, 19, 7853–7863.

(40) Suominen, H. J.; Kirrander, A. How to Observe Coherent Electron Dynamics Directly. Phys. Rev. Lett. **2014**, 112, 043002.

(41) Kirrander, A.; Saita, K.; Shalashilin, D. V. Ultrafast X-ray Scattering from Molecules. J. Chem. Theory Comp. **2016**, 12, 957–967, PMID: 26717255.

(42) Kirrander, A.; Weber, P. M. Fundamental Limits on Spatial Resolution in Ultrafast X-ray Diffraction. Appl. Science **2017**, 7, 534.

(43) Simmermacher, M.; Henriksen, N. E.; Møller, K. B.; Moreno Carrascosa, A.; Kirrander, A. Electronic Coherence in Ultrafast X-Ray Scattering from Molecular Wave Packets. Phys. Rev. Lett. **2019**, 122, 073003.

(44) Minitti, M. P.; Budarz, J. M.; Kirrander, A.; Robinson, J.; Lane, T. J.; Ratner, D.; Saita, K.; Northey, T.; Stankus, B.; Cofer-Shabica, V. et al. Toward structural femtosecond chemical dynamics: imaging chemistry in space and time. Faraday Disc. **2014**, 171, 81–91.

(45) Stankus, B.; Budarz, J. M.; Kirrander, A.; Rogers, D.; Robinson, J.; Lane, T. J.; Ratner, D.; Hastings, J.; Minitti, M. P.; Weber, P. M. Femtosecond photodissociation dynamics of 1,4-diiodobenzene by gas-phase X-ray scattering and photoelectron spectroscopy. Faraday Discuss. **2016**, 194, 525–536.

(46) Emma, P.; Akre, R.; Arthur, J.; Bionta, R.; Bostedt, C.; Bozek, J.; Brachmann, A.; Bucksbaum, P.; Coffee, R.; Decker, F. J. et al. First lasing and operation of an Ångström-wavelength free-electron laser. Nat. Phot. **2010**, 4, 641–647.

(47) Tsuneyuki, S.; Kobori, T.; Akagi, K.; Sodeyama, K.; Terakura, K.; Fukuyama, H. Molecular orbital calculation of

biomolecules with fragment molecular orbitals. Chem. Phys. Lett. **2009**, 476, 104–108.

(48) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. Phys. Chem. Chem. Phys. **2012**, 14, 7562–7577.

(49) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: An approximate computational method for large molecules. Chem. Phys. Lett. **1999**, 313, 701–706.

(50) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. Wiley Interdisciplinary Reviews: Computational Molecular Science **2014**, 4, 325–362.

(51) Besalú, E.; Carbó-Dorca, R. The General Gaussian Product Theorem. J. Math. Chem. **2011**, 49, 1769–1784.

(52) Debye, P. Zerstreuung von röntgenstrahlen. Ann. Phys. **1915**, 351, 809–823.

(53) Debye, P.; Bewilogua, L.; Ehrhardt, F. Zerstreuung von Röntgenstrahlen an einzelnen Molekülen (vorläufige Mitteilung). Phys. Zeits. **1929**, 30, 84.

(54) Chantler, C. T. Theoretical Form Factor, Attenuation and Scattering Tabulation for Z=1-92 from E=1-10 eV to E=0.4-1.0 MeV. J. Phys. Chem. Ref. Data **1995**, 24, 71.

(55) Chantler, C. T. Detailed tabulation of atomic form factors, photoelectric absorption and scattering cross section, and mass attenuation coefficients in the vicinity of absorption edges in the soft X-ray (Z=30–36, Z=60–89, E=0.1 keV–10 keV), addressing convergence issues of earlier work. J. Phys. Chem. Ref. Data **2000**, 29, 597–1056.

(56) Brown, P.; Fox, A.; Maslen, E.; O'Keefe, M.; Willis, B. Int. Tab. Cryst. Vol. C; Springer, 2006; pp 554–595.

(57) Inokuti, M. Inelastic Collisions of Fast Charged Particles with Atoms and Molecules: The Bethe Theory Revisited. Rev. Mod. Phys. **1971**, 43, 297–347.

(58) Mezey, P. G. Fuzzy Electron Density Fragments in Macromolecular Quantum Chemistry, Combinatorial Quantum Chemistry, Functional Group Analysis, and ShapeActivity Relations. Acc. Chem. Res. **2014**, 47, 2821–2827.

(59) Niebling, S.; Bjrling, A.; Westenhoff, S. MARTINI bead form factors for the analysis of time-resolved X-ray scattering of proteins. J. Appl. Cryst. **2014**, 47, 1190–1198.

(60) D. G. Fedorov, J. C. K.; Jensen, J. H. Empirical corrections and pair interaction energies in the fragment molecular orbital method. Chem. Phys. Lett. **2018**, 706, 328–333.

(61) Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Ab Initio Quantum Chemistry for Protein Structures. J. Phys. Chem. B **2012**, 116, 12501–12509.

(62) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martnez, T. J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm-Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. J. Chem. Theory Comp. **2011**, 7, 1814–1823.

(63) Rudberg, E. Difficulties in applying pure Kohn-Sham density functional theory electronic structure methods to protein molecules. J. Phys. Cond. Mat. **2012**, 24, 072202.

(64) Hameed, R.; Khan, A.; van Mourik, T. Intramolecular BSSE and dispersion affect the structure of a dipeptide conformer. Mol. Phys. **2018**, 116, 1236–1244.

(65) Ryde, U.; Olsen, L.; Nilsson, K. Quantum chemical geometry optimizations in proteins using crystallographic raw data. J. Comp. Chem. **2002**, 23, 1058–1070.

Graphical TOC Entry